'Swiss Voice App': A smartphone application for crowdsourcing Swiss German dialect data

Kolly, Marie-José ; Leemann, Adrian ; Dellwo, Volker ; Goldman, Jean-Philippe ; Hove, Ingrid ; Ibrahim, Almajai

Since it is has application in creating a critical edition, bibliography and book history research, this tool has the capability of gaining widespread adoption.

Future Work

Beyond printed material, it will be interesting to evaluate the tool for handwritten documents and make it robust for such documents. Also it will be great to test the tool for non-English documents. We can try out different visualization formats and different ways the scholars can use the output in their work. A detailed usability study can be conducted where scholars can perform some real collation work on few pages and compare their traditional method and the vHinman. Also the accuracy could be tested for warped images as most of the unobtrusive scanning methods produce some warping on the images.

References

Unsworth J. (2000) *"Scholarly Primitives: what methods do humanities researchers have in common, and how might our tools reflect this?"* In Humanities Computing: formal methods, experimental practice (13 May 2000)

Schmidt, D., & Colomb, R. (2009). *"A data structure for representing multi-version texts online."* Journal of Human–Computer Studies, 67(6), 497–514.

Smith, S.E. (2002) *"'Armadillos of Invention': A Census of Mechanical Collators, Studies in Bibliography"* 55, pp. 133-170

Cream R. *"Sapheos Project"*, CDH University Of Southern Carolina sapheos.org

Raabe W. "*Collation in Scholarly Editing: An Introduction"* wraabe.wordpress.com/2008/07/26/collation-in-scholarly-editing-an-introduction-draft

www.juxtasoftware.org

collatex.net

Lowe, D.G. (2004) *"Distinctive image features from scale-invariant keypoints."* Int. J. Comput. Vision, 60:91–110, November 2004.

Audenaert, N. and Furuta, R. (2010) *What Humanists Want: How Scholars use Primary Source Documents*, Proceedings of the 10th Annual Joint Conference on Digital Libraries, pp. 283-292.

Yalniz, I. and Manmatha, R., *"An efficient framework for searching text in noisy document images,"* Proceedings of the 10th IAPR International Workshop on Document Analysis Systems (DAS'12)

Robinson, Peter M. W. (1994). "*Collation, textual criticism, publication, and the computer."* Text 7: 77-94. www.jstor.org/stable/pdfplus/30227694.pdf. 16 Dec. 2011

Audenaert, N., Lucchese, G. and Furuta, R. *"CritSpace: a workspace for critical engagement within cultural heritage digital libraries"* ECDL'10 Proceedings of the 14th European conference on Research and advanced technology for digital libraries

Audenaert, N. *"CritSpace: An interactive visual interface to digital collections of cultural heritage material"*

Marshall, C.C., Shipman, F.M. *"Spatial hypertext and the practice of information triage*." In: ACM Conference on Hypertext (Hypertext 1997), pp. 124–133 (1997)

Rosten E., and Drummond T. (2006) *"Machine Learning for high-speed corner detection."* In European Conference on Computer Vision. Volume I, 430-443, May 2006.

Swiss Voice App: A smartphone application for crowdsourcing Swiss German dialect data

Kolly, Marie-José
University of Zurich, Switzerland

Leemann, Adrian
University of Zurich, Switzerland

Dellwo, Volker
University of Zurich, Switzerland

Goldman, Jean-Philippe
University of Geneva, Switzerland

Hove, Ingrid
University of Zurich, Switzerland

Almajai, Ibrahim
University of Geneva, Switzerland

1 Introduction

The spatial variability found in dialects is an essential indexical property that is highly salient to listeners in everyday language situations: at social events, for example, one often hears conversations of the type "I have trouble localizing your dialect – where do you come from?". Although listeners are typically unaware of the underlying linguistic mechanisms involved, they are actively engaging in perceptual dialectology (cf. Preston 1989, Clopper & Pisoni 2004) and they seem keenly aware of dialectal variation. It is interesting then that different language speaking groups seem to recognize dialects of their language with different degrees of accuracy. Leemann & Siebenhaar (2008) and Guntern (2011) show that naïve Swiss German listeners can accurately recognize a speaker's dialect with a recognition rate of 86% and 74% respectively. However, Clopper & Pisoni (2005) report identification rates of only 30–50% for American and British English dialects; Kehrein, Lameli & Purschke (2011) report similar recognition rates for German dialects. Recent studies show that dialect recognition is possible via the mobile application *Dialäkt Äpp* (Leemann & Kolly, 2013; Kolly & Leemann, in review).

This contribution describes work in progress: *Voice Äpp*, currently in development at the University of Zurich, is a follow-up project on *Dialäkt Äpp*. The main purpose of both smartphone apps is to identify users' dialects on the basis of the dialectal variants of 16 words. *Dialäkt Äpp* users provide their pronunciation through tapping on the corresponding variant on the smartphone screen. However, the new *Voice Äpp* asks users to pronounce the word and uses automatic speech recognition (ASR) to identify users' pronunciation variants. The ASR training for *Voice Äpp*is partly based on acoustic data crowdsourced through *Dialäkt Äpp.Voice Äpp*further aims at illustrating the individuality in users' voices by providing a multidimensional profile of their voice. The launch of *Voice Äpp*is planned in December 2014.

Several research teams are interested in creating similar applications for other languages, using the frameworks put forth by *Dialäkt Äpp* and *Voice Äpp*: Mobile applications that recognize regional varieties of the entire German-speaking area, of American English, of British English, and of Italian, are currently under development.

2 Crowdsourcing data with Dialäkt Äpp

In 2013 we launched the iOS application *Dialäkt Äpp*, which capitalizes on the Swiss public interest in dialectology (Leemann & Kolly, 2013). We provided a functionality that, on the one hand, allows users to localize their own Swiss German dialect by indicating their pronunciation of 16 words (see Figure 1). Given the task to predict Swiss German dialects, a model was built by phoneticians who devised a set of maximally predictive words (i.e. maps from the Linguistic Atlas of German-speaking Switzerland: *Sprachatlas der Deutschen Schweiz* (SDS, 1962–2003)) that capture dialectal differences between localities. On the other hand, users can record their own dialect and listen to recordings of other users, thus discover the Swiss dialectal landscape. Figure 1 shows three screens of the application: the choice of dialectal variants for the word *Donnerstag*'Thursday', the identified localities as a list and on a map (Bern being the best hit in this example) and the

distribution of users' recordings covering German-speaking Switzerland.

Fig. 1: Screens of Dialäkt Äpp: (1) choice of dialectal variants with buttons; (2) result provided as a choice of five best hits and their corresponding positions on a map; (3) users' recordings (one pin per locality)

Dialäkt Äpp was launched on March 22, 2013, and has been downloaded over 58'000 times (as of February 28, 2013). The data recorded by this application contains (a) (written) choices of pronunciation for 16 words by each user who localized his/her dialect and (b) audio data for the same 16 words by each user who chose to record his/her voice. For (a), the corpus contains data from over 42'000 subjects (58% males, 42% females). Most users are from the cantons (and capitals) of Zurich, Bern, Basel, Luzern, Aargau, and St. Gallen. 64% of the users' pronunciation variants still correspond to the local variant recorded by the SDS (1962–2003) in the 1940's and a large number of users report that the localization of their dialect by the application is very close to their dialectal origin. For (b), the corpus counts 38'477 recorded variants stemming from a total number of 2'633 iOS devices (which corresponds roughly to the number of speakers; 54% males, 46% females). The geographical distribution of users corresponds to that of the data presented in (a).

The data elicited by *Dialäkt Äpp* has great potential for dialectological as well as forensic phonetic research. It can be used to create new dialect maps and compare them to the maps published in the SDS (1962–2003), thus to track sound change in progress. A number of maps have already been created (for the words *Apfelüberrest*'apple core', *Bett* 'bed', *schneien*'to snow', *Tanne*'pine tree', and *tief*'low'). Preliminary analyses show that phonetic isoglosses, as illustrated in maps like *Bett*(quality of /e/) and *Tanne* (quanity of /n/) are congruent with data from the SDS (1962–2003) (Kolly & Leemann, in review). The data can also be used to compare dialects at the acoustic phonetic level: For example, preliminary results show differences in speaking rate between the Bern dialect and the Zurich dialect (Leemann, Kolly, & Dellwo, accepted). Furthermore, this corpus can be used to create population statistics for a variety of phonetic parameters, which is desirable for forensic phonetic voice comparison (cf. Nolan et al., 2009).

3 Development of Voice Äpp

Voice Äpp has two major aims:
- To use ASR techniques to localize users' dialects
- To provide users with a multidimensional profile of their voice

3.1 ASR-based dialect localization

The novelty of this new project is to use ASR techniques instead of multiple choice buttons. Some difficulties can be expected as the ASR approach is not error-free, especially through a mobile application: recording conditions may vary a lot due to the distance from the microphone, noisy environments etc. However, the high-resolution microphones of smartphones, iPhones in particular, should facilitate the ASR task. Furthermore, identifying dialects, where small variation has to be taken into account, is not the initial purpose of ASR systems; the speech recognition domain aims at normalizing such variation and at being rather dialect- or speaker-independent. In addition to this, the number of possible pronunciation variants for each word is important. For example, the word *Bett* 'bed' only counts two variants in the SDS (/bet/ and /bɛt/) whereas *Augen* 'eyes' has eleven dialectal pronunciation variants. Theß latter is highly discriminant – but the ASR task is more difficult. The algorithm will have to be modified since the voice recognition approach is not as reliable as the selection with buttons.

In order to achieve this, an ASR system is trained with two corpora: (a) the *Dialäkt Äpp* corpus described in 2 and (b) the TEVOID corpus (Dellwo, Leemann, & Kolly, 2012). Corpus (a) contains about six hours of speech of over 2'600 speakers, covering a dense net of local dialects in German-speaking Switzerland. Each recording is an isolated word from a set of 16 words. Corpus (b) contains two hours and 45 minutes of speech of 16 Zurich German speakers. Each recording is either a spontaneous or a read sentence. While the second corpus has been segmented by hand, the first one needs data preparation and verification as it was collected without control of linguistic content nor acoustic environment.

So far, encouraging results are obtained with limited training data. After ASR training with five variables from the *Dialäkt Äpp* corpus, dialect word recognition has reached accuracies of 92% (*Bett*'bed'), 90% (Kind 'child', *Apfelüberrest*'apple core'), 85% (*Tanne*'fir tree'), 79% (*fragen*'to ask'). These accuracies may increase with larger amounts of training data, which is currently being worked on.

3.2 Multidimensional voice profile and infotainment content

The second function of the *Voice Äpp* is a voice profile provided to the user. Based on a sentence recorded in their dialect, users learn about characteristics of their own voice in a playful way. A number of menus allow users to explore different aspects of speech, e.g. pitch, speech rate, articulation, auditory and visual perception.

Pitch: The fundamental frequency (f0) of the users' sentence is calculated and displayed in a histogram representing the distribution of the f0 of all the previous users.

Speech rate: The speech rate of the users' sentence is calculated and displayed in comparison to the previous users' speech rate.

Articulation: Users learn about sounds and their articulation. Upon clicking on an IPA symbol a sagittal cut is shown and the sound is played. In an interactive sagittal cut users move the position of the articulators and hear the corresponding vowel sound.

Auditory perception: Users can listen to what their sentence would sound like to a person with a hearing impairment/a cochlear implant.

Visual perception: Users are shown a video illustrating the *McGurk effect* (MacDonald & MacGurk, 1978) and the *Cocktail Party Effect* (Handel, 1989). Both effects illustrate that visual cues can be crucial for speech perception.

4 Conclusion

Voice Äpp should be as interactive as possible, allowing users to learn about the individual features of their dialect and their voice in a playful way. As shown by *Dialäkt Äpp*, a mobile application such as *Voice Äpp* is interesting for the user as well as for the researcher: by providing appealing content to the user, we gain large amounts of data. This crowdsourced data can be used to create population statistics, for example for analyses of speech prosodic features. In particular, *Voice Äpp* creates real time f0 and speaking time statistics, which represents a novelty for e.g. the field of forensic phonetics.

Acknowledgements

The project *Swiss VoiceApp – Your voice. Your identity* is funded by the Swiss National Science Foundation (SNSF); funding scheme: Agora; grant number: 145654.

References

Clopper, C.G., & D. Pisoni (2005). *Perception of dialect variation*. In: Pisoni, D., R.E. Remez (Eds.), The Handbook of Speech Perception, Oxford: Blackwell, 313–337.

Dellwo V., Leemann, A., & Kolly, M.-J. (2012). *Speaker idiosyncratic rhythmic features in the speech signal*. Proceedings of Interspeech2012. 9.-13.9.2012, Portland (OR), USA.

Ferragne, E., & Pellegrino, F. (2007). *Automatic dialect identification: A study of British English*. In: Speaker classification II. Berlin/Heidelberg, Springer: 243–257.

Guntern, M. (2011). *Erkennen von Dialekten anhand von gesprochenem Schweizerhochdeutsch*. Zeitschrift für Dialektologie und Linguistik 78/2: 155–187.

Handel, S. (1989): *Listening*. An Introduction to the perception of auditory events. MIT Press.

Kehrein, R., Lameli, A., & Purschke, C. (2010). *Stimuluseffekte und Sprachraumkonzepte*. In: Anders, C., Hundt, M., Lasch A. (Eds.). "Perceptual dialectology". Neue Wege der Dialektologie. Berlin/New York, de Gruyter: 351–384.

Leemann, A., & Kolly, M.-J. (2013). *Dialäkt Äpp*. https://itunes.apple.com/ch/app/dialakt-app/id606559705?mt=8.

Kolly, M.-J. & Leemann, A. (in review). *Dialäkt Äpp: Communicating dialectology to the public – crowdsourcing dialects from the public*. To appear in: Leemann, A., Kolly, M.-J., Schmid, S., & Dellwo, V. (Eds.). Trends in Phonetics in German-speaking Europe, Bern/Frankfurt: Peter Lang.

Leemann, A., Kolly, M.-J., & Dellwo, V. (accepted). *Crowdsourcing regional variation in speaking rate through the iOS app 'Dialäkt Äpp'*. To appear in: Proceedings of Speech Prosody 2014,20.–23.05.2014, Dublin.

Leemann, A., & Siebenhaar, B. (2008). *Perception of Dialectal Prosody.*Proceedings of Interspeech 2008.

MacDonald, John, & MacGurk, Harry (1978). *Visual influence on speech perception processes*. Perception & Psychophysics 24/3: 253–257.

Nolan, F., McDougall, K., de Jong, G., & Hudson, T. (2009). *The DyViS database: style-controlled recordings of 100 homogenous speakers for forensic phonetic research*. The International Journal of Speech, Language and the Law 16/1: 31–57.

SDS Sprachatlas der deutschen Schweiz. (1962-2003). Bern (I-VI), Basel: Francke (VII-VIII).

Beautiful lips and porcelain cheeks: extracting physical descriptions from recent Dutch fiction

Koolen, Corina
c.w.koolen@uva.nl
University of Amsterdam, Netherlands, The

Wubben, Sander
s.wubben@uvt.nl
Tilburg University, Netherlands, The

van Cranenburgh, Andreas
andreas.van.cranenburgh@huygens.knaw.nl
University of Amsterdam, Netherland

1. Introduction

In literary analysis, description – as opposed to narration – has previously often been an underestimated part of fiction. Literary theorists such as Bal, Lopes and Nünning however have made a case for its relevance [1, 6, 8]. Lopes reviews how well-known theorists like Barthes have dismissed description as 'extra', irrelevant or stalling the plot; he counters these notions with the statement that "[d]escription and narration constitute the two most basic modes of structuring any prose fiction text" [6, p. 19]. How the plot is conveyed, is relevant for how a text is judged. Literary theorist Wells for instance argues that description is the distinguishing factor between quality literature and 'simple' chick-lit novels [15]. Indeed, research has shown that literary novels contain significantly more noun phrases and prepositional phrases than chick lit, indicating a larger amount of description [5]. In this paper, the first steps are taken of a larger project in which description in fiction is computationally analyzed, as opposed to the now popular computational analysis of narrative (see for instance 7). The preliminary question that we want to answer is: how (well) can we extract descriptions from fiction? This will be tested in the current paper by zooming in on a specific domain: the physical description of fictional characters.

2. Motivation

Descriptions of physical appearance are chosen as a test case as they are more likely to occur in a current-day novel than for instance landscape description. Moreover, main characters are often introduced in the first chapters. This makes it possible in case of manual tagging (which we have done) to tag only the first chapters of a novel. Finally, it would be an interesting feature for further literary interpretation. Connotations of beauty in folk tales have been researched [i.e. 14], but this has not yet been done for novels.

3. Method

The corpus of [5] is used, consisting of 32 novels of recent Dutch fiction, half chick-lit, half literary novels. Two of them were tagged from beginning to end for descriptions of physicality, including clothing. One is a literary novel, De schilder en het meisje ('The painter and the girl') by Margriet de Moor, the other chick lit, Zwaar verliefd ('Heavily in Love') by Chantal van Gastel. Bal defines description as "a textual fragment in which features are attributed to objects" [1, p. 36], a definition we will follow. We tagged full sentences that were either mainly concerned with physical appearance (example 1a, Van Gastel), mentioned a single feature (1b, De Moor) or somewhere in between.

1a. Hij heeft mooie lippen. *He has beautiful lips.*

1b. Door de rook heen keek hij naar de porseleinen wangen van mevrouw Cloeck[.] *Through the smoke he watched madam Cloeck's porcelain cheeks[.]*

For the extraction, two approaches are compared: (1) manual development of lexical-linguistic patterns and (2) a Naive Bayes and an SVM classifier. For the former, because patterns were manually developed on the basis of two novels, the patterns were subsequently tested on the other 30 novels, each of which the first 500 sentences were manually tagged.

3.1 Lexical-linguistic patterns

After an initial exploration of the two main novels' tagged sentences, an approach was adopted of manually developing patterns to detect sentences containing description. Hearst uses similar patterns to harvest hyponyms [3]. Patterns consist of a combination of linguistic and lexical information, see example 2 below. A set of 13 patterns was written. The manual exploration showed that sentences containing physical descriptions, as opposed to sentences with no such descriptions, (a) contain more nouns and adjectives, (b) are regularly coupled with a few specific, static verbs, and (c) contain a couple of recurring base lexical-linguistic patterns, e.g., 'He was [a manNP] [[withPP] [brown eyesNP]]'. To perform extraction, the corpus was parsed with Dutch parser Alpino [2, 12]. Alpino parse trees provide rich linguistic annotations of sentences such as grammatical function of constituents. The trees can be queried with XPath, which was integrated in Van